DOCUMENT RESUME

ED 275 729 TM 860 643

AUTHOR Gallagher, Paul; And Others
TITLE Time as a Factor in Test Performance in a State Mandated Testing Program.
PUB DATE 85
NOTE 27p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Black Students; *English (Second Language); Essay Tests; Ethnic Groups; Higher Education; Hispanic Americans; Mathematics Tests; *Minimum Competency Testing; *Racial Differences; Reading Tests; Scores; *Test Bias; *Test Validity; *Timed Tests; White Students; Writing Evaluation
IDENTIFIERS *College Level Academic Skills Test; Florida

ABSTRACT

To alleviate the possibility that Florida's College Level Academic Skills Test (CLAST) is biased against non-native English speakers because of its time limits, additional time was allowed for some samples. CLAST, designed to measure communicative and computational competence in college sophomores, has four subtests: reading, writing, computation, and essay. The research design, using four groups, allowed 20 additional minutes for one subtest and 10 additional minutes for another subtest. CLAST normally requires 3 hours and 45 minutes. One sample included 598 students and another sample included 659. The racial and ethnic composition of the two samples were 63 and 55 percent Caucasian, 23 and 27 percent Hispanic, and 7 and 14 percent Black. At the end of testing, students completed a questionnaire assessing the adequacy of allowed time. The results were based on small sample sizes, especially for Blacks, and should be viewed as tentative. The low failure rate also limited the ability to define relationships among the variables. Based on findings from the two field tests, it was concluded that no significant relationships could be found between extra time, racial/ethnic characteristics, and performance on any of the subtests. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED275729

TIME AS A FACTOR IN TEST PERFORMANCE IN A STATE MANDATED TESTING PROGRAM

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

W. Younkin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Florida International University

Paul Gallagher
Barry Greenberg
Laura Melvin
William Younkin

1985

TM 860 643

OVERVIEW

Concern has been expressed by a number of institutions that the time limits of the CLAST exam may result in an unfair test administration procedure which could produce a bias against some students; especially those whose native language is not English. Several studies have examined the effects of test speediness, or time limits, of standardized tests upon performance of examinees. In terms of the impact upon minorities or other groups of interest, the results have been mixed. Evans and Reilly (1972) examined the effects of increasing the amount of testing time per item on the reading comprehension section of the Law School Admission Test (LSAT). The same authors (1973) increased the time limit on a special quantitative section of the Admission Test for Graduate Study in Business (ATGSB). They were unable to find differential improvement in test scores for minority students in either study. These results were supported by the findings of Wild, Durso, and Rubin (1982). These authors found that increased time limits on special verbal and quantitative sections of the Graduate Record Examination (GRE) did not differentially improve the scores of minority group examinees. On the other hand, Daly and Stahman (1968) found that increasing the time limits on the Cooperative English Expression Test did improve the performance of freshman college students who had been placed into remedial English classes. Flores and Seaman (1974) also found that Black Americans performed significantly better on the untimed General Education Development test (GED) than the timed version of the same test. Similar results were obtained by Immerman (1980) who found the test scores of Native American Indian students were significantly higher on the untimed version of the Stanford Diagnostic Reading Test.

DEVELOPMENT OF CLAST

As part of the overall effort to raise student achievement in the State of Florida, faculty members from community colleges and state universities worked for two years in statewide committees and agreed upon the communication and computation skills which should be expected of all students who complete the sophomore year in college. The State Board of Education has adopted those skills as college-level communication and computation skills in its Rule 6A-10.31, FAC.

DEVELOPMENT OF CLAST (Continued)

As directed by the State Legislature, the Department of Education developed a test, the College-Level Academic Skills Test (CLAST), to measure student achievement of the agreed-upon skills in community colleges and state universities. Beginning with the initial administration of CLAST in October, 1982, scores on each of the four subtests in CLAST have been required by Sections 240.233(5) and 240.239(3) and 240.319(3)FS of all students who are seeking either an Associate of Arts degree from a community college or a public university in Florida or admission to upper division status in a state university.

For any term which begins on or after August 1, 1984, student scores on CLAST must equal or exceed passing scores which have been set by Rules of the State Board of Education.

At the time the State Board of Education adopted the list of college-level communication and computation skills in October, 1981, the Board directed through Rule 6A-10.31 that steps be taken to develop a test to measure achievement of those skills. Specifically, the Board directed that specifications be written for items to measure the skills, and that a test be developed with items written to those specifications.

Pursuant to those directions, the Department of Education, through the College-Level Academic Sills Program (CLASP), undertook to develop specifications for test items to measure all of the skills that are included in Rule 6A-10.31, FAC. Item specifications have been written for all of the skills in the four skill areas to be tested, and the specifications have been issued by the Commissioner of Education for use in the community colleges and state universities of Florida.

Test items were then written for all of the skills that are measured by CLAST. The Department of Education entered into a contract with the University of South Florida for the development of objective items for the computation subtest and with the University of Florida for the development of objective test items for reading and writing skills and essay topics for writing skills. The work under those contracts involved personnel from community colleges as well as state universities. Prior to the acceptance of test items by the Department, the test items were subjected to in-depth review by content and measurement specialists as well as bias reviewers from the faculties of community colleges and state universities and personnel from the Department of Education.

Prior to development of test forms for administration beginning in October, 1982, the objective test items and essay topics were field tested. Under a contract between the Department and the Florida State University objective test items were field tested with students in twenty-one community colleges and state universities. Field test essays were written by students in three state universities and five community colleges.

DESCRIPTION OF THE PROBLEM

An important dimension of the impact of passing scores on CLAST will be derived from the likelihood that the initial effect on individuals may differ by their sex, race, and ethnic derivation and by the nature of their previous educational experiences.

It must also be recognized that the educational institutions in Florida are impacted by the fact that there are standards which are external to their own instructional processes. Institutions will be required to assure that students who enter their college credit programs have effective opportunity to achieve the skills to the level required by the passing scores. Admissions counseling, development and remedial programs, standards for course completion, and standards of student progress that are required for retention as well as enrollment and funding will be impacted by the passing scores on CLAST.

Concern has been raised that an unreasonable bias has been introduced against students whose native language is not English. For example, the report of the Miami-Dade Community College CLAST Standards Setting Panel stated that "Many students who have second-language problems do not have sufficient time to complete the composition and/or reading portion of the CLAST. For these students, and for many others, it is important that the CLAST, which is a competency-based test, be a power test, not a speed test." The proposed solution was that the time permitted for the CLAST examination should be extended for the reading and the essay tests.

The issue, then, is whether the time allotment for the subtests of CLAST has a differential effect on students who are from various ethnic backgrounds.

DESCRIPTION OF THE STUDY

Purpose of Study:

The purpose of this study was to determine whether the addition of testing time will result in significant gains in CLAST performance of students with different ethnic backgrounds. This study was conducted in two parts. The initial part was conducted in the Fall of 1984, and the second was conducted at the following scheduled CLAST administration in the Spring of 1985.

Hypotheses:

It is hypothesized that students who receive an "extra" twenty (20) minutes on a subtest of CLAST will significantly outperform students who receive an "extra" ten (10) minutes and those who have no "extra" time. Additionally, it is hypothesized that students who receive an "extra" ten minutes on a subtest of CLAST will significantly outperform students who receive no "extra" time.

DESCRIPTION OF THE STUDY (Continued)

Subjects:

The subjects for Part 1 of this study were the 598 students who were administered the CLAST during the Fall of 1984 at both campuses of Florida International University. Approximately 20% of these students were native to FIU (entered as Freshmen) while the remaining 80% were transfer students. Of the total group, approximately 63% were Caucasian, 23% are Hispanic, 7% are Black and 16% Other.

The subjects for Part 2 of this study were 659 students from the three campuses of Miami Dade Community College and both campuses of Florida International University who took the CLAST exam during the Spring administration, 1985. Of this group, approximately 55% were Caucasian, 27% Hispanic, 14% Black and 4% Other.

Description of the Instrument:

CLAST consists of four subtests - Reading, Writing, Computation, Essay. The reading subtest uses objective test items to measure eleven skills in the areas of literal comprehension and critical comprehension. The objective writing subtest measures fourteen skills in the areas of word choice, sentence structure, grammar, spelling, and punctuation. The essay writing subtest measures twenty-four writing skills. The computation subtest uses objective items to measure thirty-five skills in the areas of arithmetic, algebra, geometry and measurement, logical reasoning, and statistics including probability.

The administration of CLAST requires three hours and forty-five minutes of actual testing time. It is given in a single morning session with one twenty minute break.

Design:

The research design (Figure 1) calls for four groups. Each group received an additional twenty minutes in one subtest and an additional ten minutes in another.

	GROUP 1	GROUP 2	GROUP 3	GROUP 4
ESSAY	20	10		
WRITING	10	20		
READING			20	10
COMPUTATION			10	20

Figure 1: Research Design - Additional Time in Minutes by Group Membership.

Procedures:

All of the students received instructions to report to a main room assignment station at least one hour prior to test administration time. Assignments were made on the basis of research design. As students arrived, they were randomly assigned to a specific testing room. All sat for the CLAST in accordance with the Statewide Test Administration Plan. The only difference was that, depending on the grouping, students received additional time for two of the four subtests. The students were told that they were receiving additional time as the instructions for the subtests were given. All changes made to the oral instructions due to timing factors were typed into the standardized administration manuals and were read in a consistent manner across groups.

At the end of the testing day, the students were required to complete a brief questionnaire regarding the adequacy of the time allotted for each of the subtests.

RESULTS OF PART 1 OF THE STUDY

1. Sample description and questionnaire results:

A total of 646 students from Florida International University were placed in one of four groups according to the approach specified above and complete data were available for 587 students. Significant characteristics of the sample are described in the tables which appear in the appendix. They reveal, among other things, a satisfactory representation of the racial/ethnic balance existing in the population at large for all except Blacks. Group placement did result in adequate percentages from each sub-group receiving each "treatment." Note that for Blacks, 22.2% were placed in group 1, 31.1% in group 2, 22.2% in group 3 and 24.4% in group 4; for Hispanics, 23.6 were placed in group 1, 40.7% in group 2, and 17.9% in groups 3 and 4. (See Table 1.)

On the questionnaire which the sample completed after they finished the CLAST, they were asked whether they felt they needed more or less time on any of the subtests. A summary of the responses reveals, among other things that students from group 3 tended to want more time for the essay subtest than did the other groups (this group received extra time in reading and computation); groups 1 and 4 tended to want the most extra time for the reading subtest (group 4 received an extra 10 minutes for the reading subtest, while group 1 had the usual amount of time); on the writing test, group 2 (which received the most extra time for reading) emerged as the group least likely to say they needed more time; and on the computation subtest group 4, which received the most extra time for computation) was the least likely to indicate a need for addtional time. Examining the relationship between groups, racial/ethnic background and extra time needed revealed with respect to the principal hypotheses of this study, that Hispanics who received extra time on the essay, writing or reading subtest were not less likely to say they needed more time on this subtest than were their peers who did not receive extra time.

On the question of overall time for the total exam, a relationship by racial/ethnic background did emerge with a significantly higher percentage of Blacks than Caucasians or Hispanics indicating "not enough" time (33.6% of the former compared with 9.6% of Caucasians and 13.6% of Hispanics). Finally, a relationship did not emerge between racial/ethnic group and whether the student had attended a CLAST workshop.

	GROUP 1	GROUP 2	GROUP 3	GROUP 4	TOTAL
Caucasian	111 (64%)	141 (63%)	88 (66%)	67 (58%)	407 (63%)
Black	10 (67%)	14 (6%)	10 (8%)	11 (10%)	45 (7%)
Hispanic	33 (19%)	57 (26%)	25 (19%)	25 (22%)	140 (22%)
Other	20 (36%)	11 (20%)	11 (20%)	12 (22%)	54 (8%)
TOTAL	174 (27%)	223 (35%)	134 (21%)	115 (18%)	646 (100%)

Table 1: Group compositions by racial/ethnic background.

2. Tests of Hypotheses:

The hypotheses of this study were addressed with two kinds of analyses. First, a factorial analysis of variance was conducted in which group membership and ethnic/racial background two factors with each subtest of the CLAST serving as a dependent variable. In this procedure both the main effect for group and the interaction effect between group and ethnic/racial were to be examined. (Main effects for ethnic/racial background are not relevant to this study.) Following this analysis a chi-square was performed which examined the relationship between passing or failing each subtest and group membership and ethnic/racial background. The results are presented separately for each subtest below. (Tables 2 through 5 provide the means and standard deviations by subtest, by race, by group.)

a. Results for Computation:

The anova reveals a non-significant group effect (p .055) with the results favoring group 2 which did not receive extra time for computation. In fact group 4, which received the most time for computation did the poorest on this subtest. (The above anomalies may be due more to the differences in racial composition between groups 2 and 4 than to the differences in time allotted this subtest.) A chi-square for computation by ethnic/racial and group membership must be interpreted with caution throughout because of the low numbers from any of the groups failing any of the subtests. Nevertheless, we note that none of the Blacks or Hispanics given 20 minutes additional time for computation failed this subtest: Blacks in the other groups failed at a rate ranging from 11 to 30%, while from 3 to 12.5% of Hispanics from other groups failed this subtest. (See Tables 6 and 7.)

b. Results for Reading:

There were no significant effects for group membership and none for interaction. A chi-square revealed no advantage to Blacks given extra time for the reading subtest; indeed a higher, but non-significant, percentage failed from the group given the most time for reading than from the other groups. For Hispanics, however, while the chi-square was not significant, none of those given the extra time for reading failed while 4 to 12% from the other groups did not pass this subtest. (See Tables 8 and 9.)

c. Results for Writing:

There were no significant effects for group membership but there was a significant interaction effect. Subsequent analyses revealed that Blacks from group 3 scored significantly lower than Blacks from the other groups on this subtest but this was due to the fact that one of these students from group 3 received a score of "0" on this test. This coupled with the low "n" for Blacks from group 3 suggests that no substantive meaning be attached to the significant F

Race/Group	Mean	Std.Dev.
WHITE	321.40	32.96
GROUP 1	319.42	31.99
GROUP 2	322.68	33.25
GROUP 3	323.53	33.61
GROUP 4	318.95	33.49
BLACK	301.76	29.08
GROUP 1	320.44	28.85
GROUP 2	299.92	31.31
GROUP 3	288.88	19.14
GROUP 4	299.10	29.02
HISPANIC	314.51	32.77
GROUP 1	305.24	30.64
GROUP 2	318.76	35.04
GROUP 3	319.00	27.81
GROUP 4	312.92	33.76
OTHER	307.07	25.36
GROUP 1	311.33	29.34
GROUP 2	307.75	23.53
GROUP 3	305.85	30.64
GROUP 4	298.57	9.19

Table 2: Means and Standard Deviations for Writing by Race and Group.

Race/Group	Mean	Std.Dev.
WHITE	308.86	26.68
GROUP 1	308.74	25.89
GROUP 2	311.71	27.66
GROUP 3	310.78	26.28
GROUP 4	300.01	25.00
BLACK	281.90	21.10
GROUP 1	284.33	25.39
GROUP 2	278.00	18.53
GROUP 3	276.00	24.49
GROUP 4	291.00	15.73
HISPANIC	306.77	28.08
GROUP 1	298.36	23.21
GROUP 2	312.89	29.43
GROUP 3	303.70	27.97
GROUP 4	307.12	29.20
OTHER	300.57	21.22
GROUP 1	296.20	23.63
GROUP 2	309.00	23.95
GROUP 3	302.12	17.69
GROUP 4	298.57	17.03

Table 3: Means and Standard Deviations for Computation by Race and Group.

Race/Group	Mean	Std.Dev.
WHITE	323.82	32.10
GROUP 1	320.41	34.58
GROUP 2	325.82	31.73
GROUP 3	327.34	29.46
GROUP 4	320.30	31.94
BLACK	293.76	30.89
GROUP 1	292.22	28.88
GROUP 2	291.64	23.45
GROUP 3	289.88	38.70
GROUP 4	301.60	37.25
HISPANIC	317.13	34.82
GROUP 1	311.69	37.71
GROUP 2	322.85	34.64
GROUP 3	313.79	35.46
GROUP 4	314.72	30.57
OTHER	305.39	24.77
GROUP 1	310.40	29.21
GROUP 2	309.75	14.56
GROUP 3	297.25	27.91
GROUP 4	299.00	20.31

Table 4: Means and Standard Deviations for Reading by Race and Group.

Race/Group	Mean	Std.Dev.
WHITE	5.06	1.42
GROUP 1	5.09	1.38
GROUP 2	5.19	1.48
GROUP 3	4.98	1.47
GROUP 4	4.81	1.26
BLACK	4.26	1.69
GROUP 1	5.33	1.80
GROUP 2	4.28	1.32
GROUP 3	3.11	1.45
GROUP 4	4.30	1.82
HISPANIC	4.45	1.58
GROUP 1	4.36	1.53
GROUP 2	4.71	1.76
GROUP 3	4.00	1.35
GROUP 4	4.44	1.35
OTHER	4.47	1.50
GROUP 1	4.26	1.09
GROUP 2	5.00	1.77
GROUP 3	4.25	1.66
GROUP 4	4.57	1.90

Table 5: Means and Standard Deviations for Essay by Race and Group.

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF. OF F	
MAIN EFFECTS	34503.55	6	5750.59	8.37	.00	N.S.
VAR 14 (Group)	5247.42	3	1749.14	2.54	.05	N.S.
VAR 3 (Race)	28212.76	3	9404.25	13.70	.00	N.S.
2-WAY INTERACTIONS	7868.93	9	874.32	1.27	.24	N.S.
VAR 14 VAR 3	7868.93	9	874.32	1.27	.24	N.S.
EXPLAINED	42372.49	15	2824.83	4.11	.00	
RESIDUAL	399465.43	582	686.36			
TOTAL	441837.92	597	740.09			

Table 6: Analysis of Variance Source Table for Computation - Group and Race.

	BLACK		HISPANIC	
GROUP	FAILED	PASSED	FAILED	PASSED
1	1 (11%)	8 (89%)	1 (3%)	32 (97%)
2	3 (21%)	11 (79%)	1 (2%)	55 (98%)
3	3 (30%)	7 (70%)	3 (13%)	21 (87%)
4	0 (0%)	10 (100%)	0 (0%)	25 (100%)
TOTAL	7 (16%)	36 (84%)	5 (4%)	133 (96%)

Table 7: Pass/Fail Summary for Computations for Black and Hispanic Students.

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF. OF F	
MAIN EFFECTS	46821.23	6	7803.54	7.44	.00	N.S.
VAR 14 (Group)	3720.20	3	1240.06	1.18	.31	N.S.
VAR 3 (Race)	42012.88	3	14004.29	13.36	.00	N.S.
2-WAY INTERACTIONS	5184.02	9	576.00	.55	.83	N.S.
VAR 14 VAR 3	5184.02	9	576.00	.55	.83	N.S.
EXPLAINED	52005.25	15	3467.01	3.30	.00	
RESIDUAL	608729.71	581	1047.72			
TOTAL	660734.96	596	1108.61			

Table 8: Analysis of Variance Source Table for Reading - Group and Race.

GROUP	BLACK FAILED	BLACK PASSED	HISPANIC FAILED	HISPANIC PASSED
1	1 (11%)	8 (89%)	1 (12%)	29 (88%)
2	1 (7%)	13 (93%)	3 (5%)	53 (95%)
3	2 (20%)	8 (80%)	0 (0%)	24 (100%)
4	1 (10%)	9 (90%)	1 (4%)	24 (96%)
TOTAL	5 (12%)	38 (88%)	8 (6%)	130 (94%)

Table 9: Pass/Fail Summary for Computations for Black and Hispanic Students.

for interaction. A chi-square which was not significant, revealed that while none of the Blacks from the groups receiving extra time for writing failed this subtest, one each from the other two groups did not pass. For Hispanics the results were equally non-significant with virtually all students passing (as was the case with Blacks) and the failures more or less equally distributed among the groups. (See Tables 10 and 11.)

d. Results for Essay:

There was a significant group effect (p .03) favoring groups 1 and 2, those which received the most time for essay. There was no interaction effect. A chi-square for Blacks was not significant (p .0528) but revealed that those given the most time for essay, groups 1 and 2 had the lowest failing rate. This relationship did not hold true for Hispanics where groups 1 and 2 did only slightly better than group 3 and much worse than group 4. (See Tables 12 and 13.)

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF. OF F	
MAIN EFFECTS	23671.12	6	3945.18	3.78	.00	N.S.
VAR 14 (Group)	2143.92	3	714.64	.68	.56	N.S.
VAR 3 (Race)	20836.69	3	6945.56	6.67	.00	N.S.
2-WAY INTERACTIONS	9154.49	9	1017.16	.97	.45	N.S.
VAR 14 VAR 3	9154.49	9	1017.16	.97	.45	N.S.
EXPLAINED	32825.61	15	2188.37	2.10	.00	
RESIDUAL	605042.53	581	1041.38			
TOTAL	637868.15	596	1070.24			

Table 10: Analysis of Variance Source Table for Writing - Group and Race.

GROUP	BLACK FAILED	BLACK PASSED	HISPANIC FAILED	HISPANIC PASSED
1	0 (0%)	9 (100%)	2 (6%)	31 (94%)
2	0 (0%)	14 (100%)	1 (2%)	55 (98%)
3	1 (10%)	9 (90%)	1 (4%)	23 (96%)
4	1 (10%)	9 (90%)	2 (8%)	23 (92%)
TOTAL	2 (5%)	41 (95%)	6 (4%)	132 (96%)

Table 11: Pass/Fail Summary for Writing for Black and Hispanic Students.

SOURCE OF VARIATION	SUM OF SQUARES	DF	MEAN SQUARE	F	SIGNIF. OF F	
MAIN EFFECTS	73.82	6	12.30	5.61	.00	N.S.
VAR 14 (Group)	15.58	3	5.19	2.37	.07	N.S.
VAR 3 (Race)	58.26	3	19.42	8.86	.00	N.S.
2-WAY INTERACTIONS	25.55	9	2.83	1.29	.23	N.S.
VAR 14 VAR 3	25.55	9	2.83	1.29	.23	N.S.
EXPLAINED	99.37	15	6.62	3.02	.00	
RESIDUAL	1271.16	580	2.19			
TOTAL	1370.54	595	2.30			

Table 12: Analysis of Variance Source Table for Essay - Group and Race.

	BLACK		HISPANIC	
GROUP	FAILED	PASSED	FAILED	PASSED
1	1 (11%)	8 (89%)	7 (21%)	26 (79%)
2	2 (14%)	12 (86%)	12 (21%)	44 (79%)
3	6 (60%)	4 (40%)	6 (25%)	18 (75%)
4	3 (30%)	7 (70%)	3 (12%)	22 (88%)
TOTAL	12 (28%)	31 (72%)	28 (20%)	110 (80%)

Table 13: Pass/Fail Summary for Essay for Black and Hispanic Students.

CONCLUSIONS

While results for Hispanics are based on a satisfactory sample size (n=140; 21.7% of the sample), those for Blacks (n=45; 7.0% of the sample) are not and hence findings for this group are tentative at best. The extremely low percentage of failures on any of the subtests also limits the ability to derive any relationships among the variables considered. With these points in mind, the following appear worthy of note at this point in the investigation.

1. While the results for computation are generally in a direction which would indicate that both Hispanics and Blacks benefit from additional time, this cannot be stated with any certainty at this time. The results suggest the possibility of a greater effect for Blacks, but this will have to await further verification. It may also prove to be important that those who received the extra time in computation are least likely to indicate a need for more time.

2. The results for the reading subtest seem to hold out little promise for the existence of a significant effect here.

3. The results for the writing subtest would indicate no relationship exists in this area.

4. While for Hispanics scores on the essay subtest were not related to extra time being given, for Blacks there may in fact be a relationship between these variables.

In conclusion there is the possibility that a relationship may exist between scores on two of the subtests; extra time and racial/ethnic characteristics. Further study with a larger, more heterogeneous sample should lead to more definite results.

PART 2 OF STUDY

1. Sample description and questionnaire results:

A total of 659 students from Miami Dade Community College and Florida International University were placed in one of the four experimental groups according to the approach specified above with complete data available for 651 students; the additional 123 students comprised two control groups. Significant characteristics of the sample are described in the tables which appear in the appendix. There was a satisfactory representation of the racial/ethnic balance in the population at large and the experimental groups had adequate representations of the racial/ethnic sub-groups receiving each "treatment." The following table depicts the breakdown by Group and racial/ethnic subgroup:

	Group 1	Group 2	Group 3	Group 4	Total
Caucasian	119 (33%)	46 (13%)	127 (35%)	67 (18%)	359 (55%)
Black	23 (26%)	6 (7%)	40 (45%)	21 (22%)	90 (14%)
Hispanic	48 (27%)	15 (9%)	57 (32%)	56 (32%)	176 (27%)
Other	8 (31%)	3 (12%)	7 (27%)	8 (31%)	26 (4%)
Total	198 (31%)	70 (11%)	231 (36%)	152 (23%)	651 (100%)

TABLE 1: Group composition by racial/ethnic background.

On the questionnaire which the sample completed after they finished the CLAST, students were asked whether or not English was their native language. Although Caucasians represented 56% of the total population and 55% of the experimental population, they represented 72% of the native English speakers. Hispanics, who represented 26% of the total population and 27% of the experimental population only represented 9.3% of the native English speakers. Blacks were represented proportionally in the total and experimental populations. Of the 305 students indicating native language other than English, Spanish is the native language of 67% of the students who identified themselves as Hispanic and 32% of the students who identified themselves as Caucasian. When students were asked if they used English in thinking about the answers to the test, only 23 indicated that they did not. When asked if English is used in the student's home, 175 indicated that some other language was more frequently used. These students were distributed evenly across groups and racial/ethnic subgroups.

When asked to identify the easiest and hardest tests, students rated the writing test as easiest, followed by computation with reading and essay tied for 3rd. They perceived the computation test as hardest, followed by essay, reading and writing. In view of the fact that the greatest number of failures was in the essay (113), followed by writing (69), computation (50) and reading (31), these perceptions are very interesting.

Students were asked whether they needed more or less time for each of the subsections of the test. Although Group 1 which received 20 minutes more on the essay was most likely to say they had sufficient time, Group 3 tended to feel they needed more time. On the Reading test in which Group 3 had 20 additional minutes and group 4 ten more minutes, Group 4 was most likely to feel they needed more time, with Group 1 least likely to require more time. All groups were equally satisfied with the time allotted for the Writing Test, even though Group 1 had an additional 10 minutes and Group 2 an additional 20 minutes. This high level of overall satisfaction may relate to the perception that the Writing Test was the easiest of the four. Students were generally satisfied with the amount of time they had for the Computation test although Group 3, which received an additional 10 minutes was most likely to say they needed more time. Overall, all four groups expressed a high level of satisfaction with the amount of time allotted, with about 18% of the students saying they needed more time. Group 4, however, tended to be most satisfied with the time given. Using the racial/ethnic breakdowns, Blacks tended to be least satisfied with the amount of time given, followed by Hispanics, but Whites were satisfied slightly more than their representation in the population. These results were significant ($p = .009$)

A significant relationship also emerged between racial/ethnic group and whether the students had attended a CLAST workshop. Whites were less likely to have attended, while Hispanics and "Others" were slightly more likely to have attended.

2. Test of Hypotheses:

The hypotheses of this study were addressed with two kinds of analyses. First a factorial analysis of variance was conducted in which group membership and ethnic/racial background were considered as two factors with each subtest of the CLAST serving as a dependent variable. In this procedure both the main effect for group and the interaction effect between group and ethnic/racial background were to be examined. (Main effects for ethnic/racial background are not relevant to this study.) Then a chi-square was performed which examined the relationship between pass or failure one each subtest and group membership and ethnic/racial background. The results are presented below separately for each subtest.

a. Results for Computation:

There were no significant main effects for group membership nor were there interaction effects for group and racial/ethnic background on the Anova. Although group 3 received 10 minutes more time, they did poorest on the test while Group 2 with no extra time performed the best. (see Table 2 below) Failure rates in the four groups fell between 6-8% regardless of whether they were given more time or not. No significant results were found on the chi-squares for computation.

	Whites	Blacks	Hispanics	Others
Group 1	309/35	289/25	300/26	302/27
Group 2	318/36	291/30	296/19	292/50
Group 3	307/32	290/29	303/26	280/18
Group 4	307/33	293/33	307/36	310/35

TABLE 2: Computation means and standard deviations by group and racial/ethnic background.

Thus it may be seen that the additional 10 minutes Group 3 had and the additional 20 minutes for Group 4 for Computation did not materially change the means for the subgroups. Although Hispanics seemed to do somewhat better with more time their overall passing rates were not effected. The failure rate for Blacks in the four groups was 4%, 16%, 10% and 19% for groups one through four respectively so that being given more time does not appear to have a useful effect. For Hispanics the failure rates in groups one through four were 4%, 0%, 4%, and 5% which would also argue against more time in Computation.

b. Results for Reading:

Again there were no significant results on the Anova for group or interaction effect. Group 1 did best on this test followed by Group 2 despite the fact that these two groups did not receive extra time. Significant results for Group 1 on the chi-square are probably due to the higher failure rate of Blacks in this group and the fact that none of the "others" in this group failed this subtest. Failure rates for the Blacks on the Reading test according to group were 17%, 17%, 8%, and 14%; for Hispanics, failure rates by group were 2%, 0%, 4%, and 4%. Since Group 3 received 20 minutes more time and Group 4 had 10 minutes more, it would appear that more time had no effect except possibly in the case of Blacks in Group 3. It is interesting to remember that when asked if more time was needed on the Reading Test, Group 4 was most likely to say it was required. Group means and standard deviations

are provided in Table 3 below:

	Whites	Blacks	Hispanics	Others
Group 1	323/32	294/26	314/24	300/23
Group 2	322/31	310/35	307/23	283/60
Group 3	320/29	295/29	312/22	289/28
Group 4	319/30	301/29	308/29	297/22

TABLE 3: Reading means and standard deviations by group and racial/ethnic background.

It may be seen that whereas means and standard deviations remained the same for Whites, the means for Blacks were highest in Group 2, which received no extra time and for Hispanics in Group 1 which received the standard time. Therefore it would appear that additional time is not helpful in the Reading subtest.

c. Results for Writing:

There were no significant effects for group or interaction on the anova for writing. Failure rates for Blacks and Hispanics in Writing were as follows: Blacks by group: 18%, 0%, 20%, and 14%; Hispanics, 2%, 20%, 11%, and 13%. It is interesting that the second highest failure rate for Blacks occurred in Group 1 which received 10 extra minutes and the highest failure rate for Hispanics was in Group 2 which received 20 extra minutes to complete the subtest. It is also interesting to note that there was only a 2% failure rate for Hispanics in Group 1 with 10 additional minutes. The only significant chi-square was in Group 3, which may be attributed to the high failure rate of Blacks and Others in this group. Means and standard deviations by group are presented below:

	Whites	Blacks	Hispanics	Others
Group 1	313/32	292/24	297/19	311/36
Group 2	312/30	298/18	300/31	288/65
Group 3	310/34	285/26	297/27	281/19
Group 4	312/31	294/25	295/29	294/35

TABLE 4: Writing means and standard deviations by group and racial/ethnic background.

It is interesting to note that overall means were highest for both Blacks and Hispanics in Group 2 which received an extra 20 minutes.

d. Results for essay:

In the Anova for the essay subtest, there was a significant (p .002) for group favoring Groups 1 and 2 which received more time. Group 3 had the poorest performance on this test which is the probable reason for significant results. No interaction effect for group and racial/ethnic background was significant. The failure rate for Blacks and Hispanics in the four groups are as follows: Blacks, 13%, 33%, 35%, and 14%; Hispanics, 16%, 27%, 19% and 21%. Thus the lowest failure rate for both Blacks and Hispanics was in Group 1 which had 20 minutes more time, but Group 2 with 10 minutes more time had the worst failure rate for Hispanics and second worst for Blacks. Again the chi-squares showed significant results (p .002) for Group 3 which is probably attributable to the higher failure rates for Blacks and Others in Group 3. The means and standard deviations are presented below:

	Whites	Blacks	Hispanics	Others
Group 1	5.2/1.6	4.6/1.5	4.5/1.4	5.1/1.8
Group 2	5.2/1.2	4.0/1.8	4.5/1.8	5.0/2.6
Group 3	4.7/1.5	3.8/1.5	4.6/1.3	3.1/1.6
Group 4	5.0/1.5	4.5/1.5	4.3/1.4	4.9/2.3

TABLE 5: Essay means and standard deviations by group and racial/ethnic background.

Although Blacks did best in Group 1 with 20 minutes more time, their results do not differ significantly from Blacks in Group 4 with no extra time. Hispanics in Group 3 did best with no extra time.

Conclusions For Field Test #2

Results for Hispanics are based on a satisfactory sample size (n=176; 27% of the sample), as well as a satisfactory sample for Blacks (n=90; 14% of the sample). The low percentage of failures on any of the subtests limits our ability to derive some relationships among the variables considered. With this in mind, the following conclusions are reached:

1. In computation, no significant advantage is to be found through providing students additional time. One of the groups given more time did the poorest while the group with the regular amount of time did the best.

2. In reading, no significant advantage was found related to extra time. Blacks and Hispanics did best in the groups that had no extra time on this subtest.

3. In writing there was no significant advantage related to extra time. Although there were no failures for Blacks in the group with 20 minutes extra, Blacks in the group with 10 minutes extra performed 2nd worst of all the groups. Similarly, although the lowest failure rate for Hispanics was in the group with 10 minutes extra time, the highest failure rate for Hispanics was in the group with 20 minutes extra.

4. Although there was a significant difference for group on the essay (p .002), in the group having 10 extra minutes the failure rate for Hispanics was highest of all the group conditions. The means and failure rates for Blacks in group 1 with 20 extra minutes and group 4 with no extra time are very similar. Therefore it is concluded that no advantage is to be derived by providing extra time.

In the previous study using this design, it was noted that extra time might be helpful for both Blacks and Hispanics. These findings are contradicted by the findings of this study. Results from the previous study in reading are confirmed by this study in that there seems little promise of a significant effect related to extra time on this subtest. No results were found for the writing subtest either in this study or the previous one. Although significant results on the Anova for essay were found in this study, these results are contradicted by the means and failure rates on this subtest. This tends to support the findings from the first study.

Based on these findings, it is concluded that no significant relationships can be found between extra time, racial/ethnic characteristics and performance on any of the subtests.

Appendix A

Validity and Reliability of CLAST

The College-Level Academic Skills Test has been designed to measure the achievement by college sophomores of specified communication skills and computation skills. These skills are listed in Rule 6A-10.31. While it is presumed that scores on CLAST will relate to other measures of success in college, both prior to and subsequent to the examination, CLAST does not purport to predict grade point averages or other measures of success. Rather, validity of the test rests upon the extent to which CLAST adequately measures the specific skills which it was designed to measure; that is, the extent to which the content of the test matches the set of skills. The validity of the test, then, was determined by examining the procedures for developing and selecting items for each form of CLAST.

The general procedures followed in developing the test are outlined below:

1. General test specifications, consistent with the purpose of CLAST, were developed by the CLASP staff and faculty who have expertise in both the content areas (reading, writing, and computation) and in testing.

2. Item specifications, which detail both the content and format of the items which can be developed to measure each of the skills, were developed by CLASP staff and faculty with expertise in both the content areas and testing.

3. Test items were written according to the guidelines provided by the item specifications and were reviewed by faculty and staff with careful attention given to content, measurement and bias issues.

4. Test items were field tested in community colleges and state universities.

5. Items were, for the ongoing development of alternate forms, analyzed statistically and selected for use in the test only if they meet criteria established by the CLASP staff and testing consultants.

6. The Rasch procedure was used for generating scale scores and for equating scale scores on alternate forms.

To summarize, validity of the test as a measure of achievement of specified skills, was ensured by a plan for developing and selecting items which was logically and technically sound. Content and testing specialists judge the adequacy of the items for measuring the skills, and the plan for selecting items and equating tests ensures that each form of CLAST is representative of the domain of skills being tested. Scores on each of the subtests, then, can be interpreted to be valid measures of the students' achievement of the communication and computation skills which are measured by CLAST.

The State-administered version of the CLAST was used to collect the performance data. Test reliability estimates are provided below:

Subtest	Number of Test Items	KR-20	SEM
Computation	49	.90	2.84
Reading	36	.87	2.29
Writing	29	.72	1.81
		Coefficient Alpha	
Essay		80	